

基于多层次结构建模的药物相互作用类型预测

吴怀瑾¹, 卜家梓¹, 杨念祖¹, 孙焱², 宋海涛^{1,3}, 柳宁¹, 严骏驰^{1,4*}

1. 上海交通大学人工智能教育部重点实验室, 上海 200240

2. 之江实验室, 杭州 311121

3. 上海人工智能研究院, 上海 200240

4. 上海交通大学人工智能学院, 上海 200240

* 通信作者. E-mail: yanjunchi@sjtu.edu.cn

收稿日期: 2024-04-05; 修回日期: 2024-10-03; 接受日期: 2025-01-06; 网络出版日期: 2025-01-23

国家自然科学基金重大研究计划 (批准号: 92370201) 和国家自然科学基金优秀青年基金 (批准号: 62222607) 资助项目

摘要 药物相互作用 (drug-drug interactions, DDIs) 通常在多种药物共同使用时发生. 某些药物之间的相互作用可以带来积极的治疗效果, 而某些药物相互作用引起的反应可能会带来严重的影响, 甚至危及生命, 这使得预测某两种药物之间是否存在相互作用乃至何种类型变得至关重要. 而实验研究 DDIs 的需求和成本不断增长, 凸显了采用计算方法, 尤其是机器学习方法以预测 DDIs 的价值. 然而, 此前基于机器学习的方法均未能全面建模 DDIs 问题中包括药物自身分子子结构、全结构和药物相互作用网络结构在内的 3 个层级信息, 部分方法还依赖于大规模生物知识图谱, 增加了模型的推理成本. 本文提出 HierDDI, 它仅通过数据集中的 DDIs 即能构建一个多层次图结构, 而无需依赖附加知识图谱. 更具体地说, HierDDI 通过建模药物自身分子子结构、全结构、药物相互作用网络结构这 3 个层面, 来预测药物相互作用. 由此不仅捕获了药物自身的结构信息, 还将药物间的相互作用考虑在内, 并通过端到端的训练, 建立模型对 DDIs 的全面理解. 本文所有模块的设计都非常简洁, 广泛的实验展示了本文提出的方法明显优于目前药物相互作用领域的领先方法, 成为目前表现最佳的模型. 在 DrugBank 和 Twosides 等公开数据集上, 本文提出的方法在直推式 (transductive) 场景中的准确率比当前领先方法提高了 0.58% 和 2.02%, 而在归纳式 (inductive) 场景中则更是提高了 4.20% 和 3.52%. 源代码发布在 <https://github.com/wuhuaijin/HierDDI>.

关键词 机器学习, 药物相互作用, 多层次, 图神经网络, 端到端

1 引言

随着人工智能 (artificial intelligence, AI) 技术的飞速发展, 其在药物研发中的应用 (AI4Drug) 引人注目, 正逐渐改变传统的药物发现模式. AI4Drug 结合了机器学习、深度学习等先进的计算技术, 能够显著提高药物研发的效率和精确度, 推动从药物设计到临床应用的各个阶段. 当前, AI 在药物研发

引用格式: 吴怀瑾, 卜家梓, 杨念祖, 等. 基于多层次结构建模的药物相互作用类型预测. 中国科学: 信息科学, 2025, 55: 356-371, doi: 10.1360/SSI-2024-0106

Wu H J, Bu J Z, Yang N Z, et al. Drug-drug interaction prediction via hierarchical structure modeling. *Sci Sin Inform*, 2025, 55: 356-371, doi: 10.1360/SSI-2024-0106

表 1 对比不同方法在对药物相互作用的层级结构建模方面的差异, 包括它们是否需要知识图谱以及是否为端到端训练.

Table 1 Comparison of DDI structure modeling in different methods, including whether they require KG and whether they are trained end-to-end.

Method	Substructure	Molecule	DDI-network	Require KG?	End-to-end?
BRSNMF ^[18]	✗	✗	✓	✗	✓
KGNN ^[20]	✗	✗	✓	✓	✓
MUFFIN ^[27]	✗	✓	✗	✓	✗
SSI-DDI ^[21]	✓	✓	✗	✗	✓
GMPNN ^[30]	✓	✓	✗	✗	✓
MSAN ^[29]	✓	✓	✗	✗	✓
DSN-DDI ^[31]	✓	✓	✗	✗	✓
HierDDI (ours)	✓	✓	✓	✗	✓

中的主要应用方向涵盖了分子生成^[1~3]、蛋白质设计^[4,5]、复合物结构预测^[6,7]、逆合成预测^[8,9], 以及本文所关注的药物相互作用等领域.

药物相互作用 (drug-drug interactions, DDIs) 是指药物同时或先后服用后, 在机体内发生的药理学或药效学方面的相互作用, 涉及药物的吸收、分布、代谢、排泄、药效学及药物基因组学等^[10,11]. 合理的药物组合能够增强疗效或减少不良反应, 而不当的药物组合可能导致疗效下降、毒性增加, 甚至可能引发异常反应、干扰治疗效果、加重病情, 并带来未知的风险^[12]. 多项临床数据表明不良药物相互作用会导致严重的患者不良反应事件^[13,14], 已成为药物撤市的一个重要原因. 仍有大量潜在药物相互作用待发现, 理解、预测其存在与否及类型, 对保障患者福祉、提升药效、指导药物研发与处方开具意义重大. 传统依赖实验室测试的研究方法既耗时又昂贵. 近年来药物数量增多, 对相关研究需求也在增长, 凸显出开发自动化计算工具, 尤其是基于机器学习方法预测药物相互作用的重要性. 目前, 机器学习方法已广泛用于这类潜在药物相互作用的预测任务^[15~17].

一类特定的 DDI 预测方法专注于从现有的药物相互作用网络 (简称为 DDI 网络)^[18,19] 或含有丰富生物医学信息的知识图谱 (knowledge graph, KG)^[20] 中学习药物特征, 而不考虑药物分子自身结构信息. 这些方法利用已知的 DDI 网络和大量附加信息, 预测特定一对药物之间的相互作用是否存在. 然而当已知的 DDI 网络高度稀疏, 或在极端情况下, 某些新药作为与现有药物无已知连接的孤立节点存在时, 这些方法难以进行可靠的预测. 这种场景被定义为归纳式 (inductive)^[21], 意味着要预测的药物相对于已知的 DDI 网络是新颖和未知的. 另一类 DDI 预测方法专注于理解药物的内在特性, 以预测它们之间的相互作用. 这些方法通过人工设计的特征^[22,23], 或应用神经网络学习 SMILES 序列^[24,25] 和分子图^[26,27] 特征来实现. 近年主流研究集中在用图神经网络学习分子图. 其中, 一些研究^[28,29] 强调了分子子结构与其特性存在紧密联系, 并将这些发现作为归纳偏差整合进模型中. 研究分子子结构有助于增强 DDI 预测结果的可解释性, 并在归纳式场景中增强模型泛化能力.

为了能综合上述两类方法的优势, 本文提出了一种新颖的方法从多层次视角学习 DDIs, 可以同时有效地从分子图中提取分子子结构级信息和分子全结构级信息, 并合理利用已知 DDI 网络派生的拓扑结构信息. 不同方法的比较如表 1^[18,20,21,27,29~31] 所示. 在处理具有多层次和多粒度特性的建模问题时, 如何高效提取并整合不同层级的信息成为关键挑战, 特别是实现分子图信息与 DDI 网络信息的有机结合, 以提高模型的整体性能. 此外, 本文还探讨了如何在一个统一框架下进行多层次建模, 以确保能够进行高效的端到端训练和推理. 实际上, 层次建模已在各种生物学领域得到广泛利用, 如疾病生物学^[32,33], 蛋白质组学^[34] 和神经生物学^[35~37], 本文将多层次建模引入药物相互作用领域, 并首次实现了分子子结构级、分子全结构级和 DDI 网络级的建模. 同时, 本文提出的方法并不依赖附加知识图谱, 仅在已知的 DDIs 中就能有效实现多层次建模. 图 1 展示了本文提出的多个层级结构以及与附加

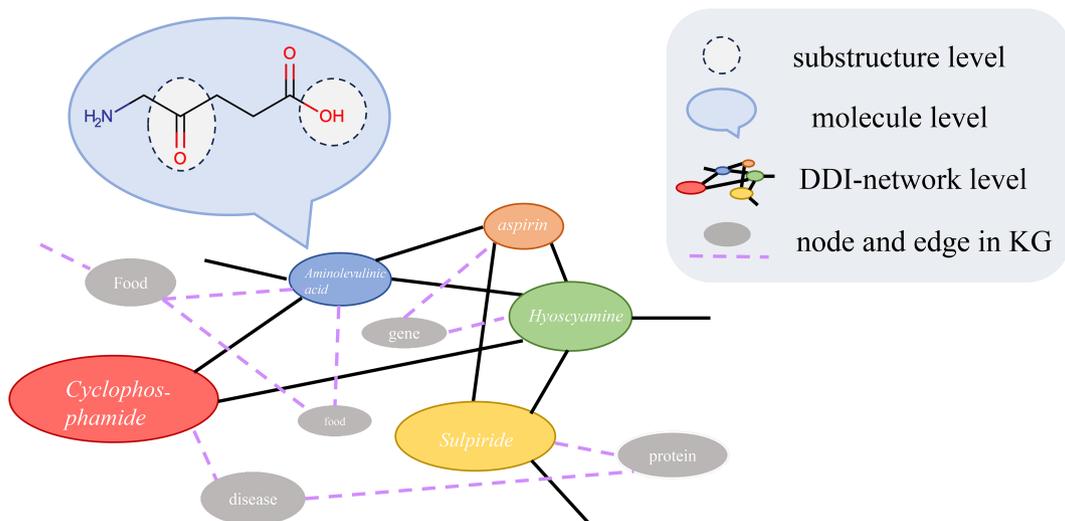


图 1 (网络版彩图) 分子子结构、分子全结构、药物相互作用网络结构等 3 个层级示意图, 以及可以额外使用的生物知识图谱。

Figure 1 (Color online) Overview of the three levels: molecule substructure, molecule structure, and the network structure of drug-drug interactions, along with the extra optional biological knowledge graph.

知识图谱的关系。

本文提出了 HierDDI。具体来说, HierDDI 使用消息传递神经网络 (message passing neural network, MPNN) [38] 提取原子级特征, 结合注意力机制捕获分子图中的重要分子子结构级特征, 并将原子特征平均聚合得到分子全结构级特征。这些药物特征随后由图神经网络 (graph neural network, GNN) 在 DDI 网络中传播, 该网络完全基于已知的 DDI 构建, 无需额外知识图谱。任意药物对之间相互作用类型的最终预测结果来自两个部分: DDI 网络更新前的特征以及 DDI 网络更新后的特征。本文提出的模型使用端到端训练范式, 多个层次信息的整合可以有效地贡献于最终结果。即使在归纳场景中, 其目标是对未知药物进行预测, 利用子结构互动模块增强了泛化能力, 使得预测性能在更广泛的药物相互作用范围内得到了提升。此外, 针对归纳式场景, 本文通过引入额外的随机边连接未知药物与已知药物, 来利用现有的 DDI 网络结构。实验结果表明, HierDDI 在两个真实公开数据集中均达到了最佳性能。作者认为, 本文提出的 HierDDI 具有以下亮点。

- 本文首次提出了一种融合分子子结构信息、分子全结构级信息及 DDI 网络级信息的图神经网络模型, 用于预测药物间的相互作用。这种多层次的建模方式能够在不同层次上全面理解和分析药物相互作用的机制和特征。如表 1 所示, 目前尚无现有工作覆盖所有这 3 个结构层次, 因此本文提出的多层次模型在药物相互作用预测领域具有独特的创新性和前瞻性。

- 本文提出的方法具备高度的通用性与灵活性, 通过端到端训练范式, 提升了 3 个层次的整体效能。该框架不仅能够整合现有研究中的模块 [28, 30], 还具有扩展新模块的潜力。

- 在两个公开数据集上的实验结果验证了 HierDDI 作为简单且强大的基线, 在直推式和归纳式任务中均取得最优性能。

2 相关研究工作

本文将从分子表示学习和药物相互作用两个方面来介绍相关研究工作。

分子表示学习。 目前分子表示学习方法主要分为两大类。第一类是基于分子 SMILES 序列的方法, 其中 SMILES 代表简化的分子输入线条条目系统 (simplified molecular input line entry system) [39]。这些方法大多采用语言模型, 如 BERT [40], 来处理分子的文本表示。SMILES 是一种用于分子的线性编

码系统, 严重依赖于遍历分子图的顺序, 这可能会限制其在诸如 DDI 预测等任务中的表达能力, 而在这些任务中分子结构提取至关重要. 第二类是基于分子结构的方法, 可以进一步分为基于分子指纹和基于图神经网络 (GNN). 分子指纹技术可以追溯到 Morgan 指纹^[41]. 然而这些基于指纹的方法通常是人工设计的, 而非为端到端训练而设计^[42]. 相比之下, 由于分子图天然存在图结构, GNNs 在分子表示学习方面越来越受欢迎. 这些基于 GNN 的方法利用分子内在图结构, 逐渐成为一类更强大的分子表示学习方法^[38, 43~46].

药物相互作用. 深度学习在药物相互作用预测问题中取得了成功, 当前主流方法基于图学习. 药物相互作用预测方法可以大致分为两类. 第一类将已知药物相互作用关系建模为 DDI 网络, 并将预测任务建模为链接预测. 这些方法基于底层网络结构预测药物之间的相互作用是否存在, 使用不同的图方法, 包括矩阵分解^[18]、图自动编码器^[47] 和对比学习^[48]. 其中一些还涉及具有丰富附加信息的生物医学知识图谱^[20, 49, 50]. 第二类利用药物的内在分子结构, 并将每对药物相互作用视为独立的正样本^[29, 51], 这意味着它们不考虑被预测的药物对与 DDI 网络中其他药物之间的相关性. 因此, 它们的重点是如何从分子结构中提取有效特征, 这与分子表示学习 (molecular representation learning, MRL) 领域密切相关, 且已在上文中介绍. 在这一类研究中, 还存在一些研究^[29, 30] 探索并强调了分子子结构建模对预测 DDIs 的重要性. 此外, 一个名为 MUFFIN^[27] 的方法结合了前述两种方法各自的一些优势. 具体而言, 它使用了专为知识图谱设计的方法 (如 TransE^[52]) 预先学习知识图谱中每种药物的表示, 并将其与从分子图中学习的药物表示结合起来用于后续预测. 这种方法不遵循完全端到端的训练范式, 限制了分子图学习和知识图谱学习之间的相互受益. 最近另一个方法 DSN-DDI^[31], 也提出两个层级的建模结构, 从单个药物和药物对中学习药物子结构. 尽管它实现了双视角建模, 但它并没有考虑到 DDI 网络, 依旧存在欠缺. 本文提出的 HierDDI 不需要依赖附加的生物医学知识图谱. 它仅基于已知的 DDI 关系, 通过分子子结构级、分子全结构级和 DDI 网络级的建模来预测 DDIs. 本文提出的方法是完全端到端训练的, 使得 3 个层级间的学习有机结合, 建立了模型对药物相互作用的全面理解.

3 多层次的药物相互作用预测模型设计

本节将公式化定义 DDI 预测问题, 再详细阐述本文提出的方法及其实现细节.

3.1 问题定义

早期的研究仅预测两种给定药物之间是否存在任何相互作用, 而不考虑具体的相互作用类型^[19, 20]. 与更近期的工作^[21, 29] 一致, 本文将采用一个更具体的设置, 即考虑相互作用类型. 下文首先介绍药物结构、DDI 网络, 然后给出 DDI 预测任务的公式化定义.

药物结构. 药物本身具有分子结构, 本文中将从分子全结构和分子子结构两个层级来提取药物自身结构信息. 在基于分子图的机器学习方法中, 分子全结构特征一般为所有原子节点特征的聚合. 而分子子结构一般指的是分子中具有某种特性的一部分结构 (例如亲水性的羟基等), 在机器学习方法中, 这些特性一般由模型自主学习得到^[28].

DDI 网络. 本文将药物相互作用形成的关系网络简称为 DDI 网络. 给定一组已知的药物相互作用关系, 可以构建一个相互作用矩阵. 形式上, 可以使用一个二进制对称邻接矩阵 $\mathbf{A} \in \{0, 1\}^{N_d \times N_d}$ 来表示 DDI 网络, 其中 N_d 表示药物数量. $\mathbf{A}_{ij} = 1$ 表示药物 d_i 和 d_j 之间存在某种相互作用, 而 $\mathbf{A}_{ij} = 0$ 表示尚未观察到它们之间的任何相互作用.

DDI 预测. 本文将药物相互作用预测简称为 DDI 预测. 给定一组已知的药物相互作用数据 $\mathcal{M} = \{(d_i, t, d_j)_k\}_{k=1}^N \subset \mathcal{D} \times \mathcal{T} \times \mathcal{D}$, 其中 \mathcal{D} 为药物集合, \mathcal{T} 为相互作用类型集合, 而目标是预测两种药物是否存在某种特定类型的相互作用. 这里将学习一个由 θ 参数化的预测模型 $\Gamma_\theta(d_i, d_j, t | \mathcal{M}, \mathbf{A})$, 其输出一个值 $\hat{y}_{ijt} \in [0, 1]$, 对应于药物 d_i 和 d_j 之间发生类型为 t 的相互作用的概率. 注意, 这里 t 的类

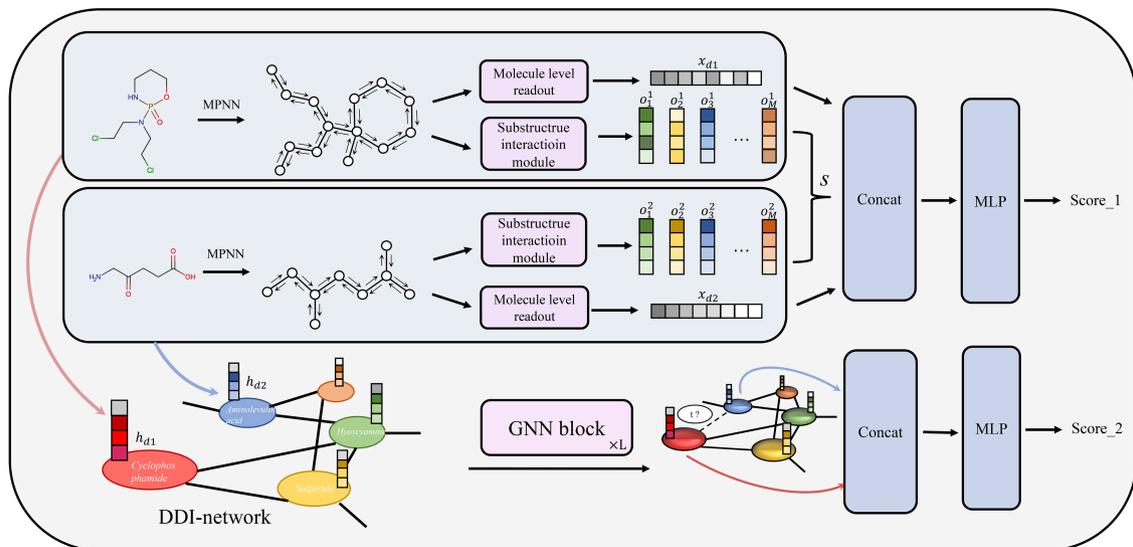


图 2 (网络版彩图) 所提出的 HierDDI 的模型架构. 药物被表示为去氢分子图, 其中节点代表原子. 分子图中原子节点的特征通过消息传递神经网络学习得到, 然后通过读出操作得到整个分子的表示 \boldsymbol{x} , 并通过子结构相互作用模块得到分子子结构的特征表示向量 $\boldsymbol{o}_1, \boldsymbol{o}_2, \dots, \boldsymbol{o}_M$, 这两者分别对应于分子全结构层级和分子子结构层级特征. 其中分子子结构提取可以参考诸如文献 [29] 的现有方法. HierDDI 基于上述信息首先预测 Score₁. 另一边, 先前获得的分子全结构层级和分子子结构层级特征用于初始化 DDI 网络中的药物节点特征, 然后经过 L 层 GNN 更新 DDI 网络中的节点特征. 最终输出的药物节点特征用于预测 Score₂. 模型将 Score₁ 和 Score₂ 的平均值作为最终预测结果.

Figure 2 (Color online) Architecture of the proposed HierDDI. Drugs are represented as hydrogen-depleted molecular graphs where each node is one atom. Atom features are learned by message passing neural networks. Then, the representation \boldsymbol{x} of the entire molecule is obtained by a READOUT operation and the representation vectors $\{\boldsymbol{o}_1, \boldsymbol{o}_2, \dots, \boldsymbol{o}_M\}$ of the substructure patterns are obtained by the substructure interaction module, corresponding to the molecule-level and substructure-level views, respectively. The substructures can be obtained by existing methods such as [29]. Then, we predict Score₁ based on the aforementioned information. Afterward, the previously obtained molecule-level and substructure-level features are used to initialize the node features in the DDI network, and the node features in the DDI network are updated by L layers of GNN. The final output node features are used to predict Score₂. We take the average of Score₁ and Score₂ as the final predicted probability.

型多样化, 包括药物之间的促进作用或是不良作用, 而标签 1 指的是存在相互作用而非正向相互作用, 标签 0 指的是不存在相互作用而非负向相互作用.

HierDDI 的模型架构如图 2 所示, 由分子表示学习模块、子结构相互作用模块、DDI 网络学习模块和预测模块组成.

分子表示学习模块. 预测药物相互作用首先需要提取药物特征. 相比 SMILES 序列, 分子图包含更丰富的信息, 且图神经网络 (GNN) 已相对成熟. 因此, 本文将分子建模为图结构, 并利用 GNN 学习其表示. 本文将药物分子表示为去氢的分子图, 并从它们的图结构中提取有效信息. 本文使用 $\boldsymbol{X} \in \mathbb{R}^{n \times \dim}$ 来表示某一分子图的原子节点特征矩阵, $\boldsymbol{E} \in \mathbb{R}^{n \times n \times b}$ 表示此分子图的化学键矩阵, 其中 n , \dim 和 b 分别是这个分子中的原子数、特征维度和化学键的类型数. 请注意, 若两个原子间没有化学键, 这也是一种特殊的类型 (即空键). 为了构建初始原子节点特征矩阵 \boldsymbol{X}^0 , 本文考虑了 8 个与原子相关的属性: 符号、度、价、电荷、电子数、杂化、氢键和芳香性. 所有这些属性都可以通过 RDKit [53] 工具包获得. 然后, 本文采用消息传递神经网络 (MPNN) [38] 来学习分子图中原子节点的特征, 该网络从中心节点的邻域聚合信息, 并迭代地更新中心节点的表示. 假设节点 p 为中心节点, 这个过程可以描述如下:

$$\boldsymbol{m}_p^k = \sum_{q \in N(p)} \text{AGGREGATE}(\boldsymbol{x}_p^{k-1}, \boldsymbol{x}_q^{k-1}, \boldsymbol{e}_{pq}), \quad (1)$$

$$\mathbf{x}_p^k = \text{UPDATE}(\mathbf{x}_p^{k-1}, \mathbf{m}_p^k), k \in \{1, \dots, K\}, \quad (2)$$

其中 AGGREGATE 使用平均聚合方式, $m_p^k = \frac{1}{|N(p)|} \sum_{q \in N(p)} m_{pq}^{k-1}$, UPDATE 使用一个神经网络 U_t (例如多层感知机) 作为更新函数. $N(p)$ 表示分子图中节点 p 的邻居, \mathbf{x}_p^k 表示第 k 次迭代后节点 p 的特征, e_{pq} 表示边 (p, q) 的特征. 经过 K 次消息传递后, 可以得到最终的原子节点特征矩阵 $\mathbf{X} = [\mathbf{x}_1^K, \mathbf{x}_2^K, \dots, \mathbf{x}_n^K]$. 然后, 对节点特征执行平均池化操作, 以获得固定维度的分子全结构级表示 $\mathbf{x} \in \mathbb{R}^{\text{dim}}$, 即 $\mathbf{x} = \frac{1}{n} \sum_p (\mathbf{x}_p^K)$. 这一步骤通常在图神经网络文献中被称为读出 (READOUT) 操作.

子结构相互作用模块. 分子子结构在药物相互作用预测中至关重要, 因为它能够揭示药物间的局部相互作用模式, 从而提供更精细的特征用于模型的准确性提升. 为了捕捉分子子结构之间的相互作用, 受文献 [29] 启发, 本文采用了类 Transformer 架构 [54] 从分子图中提取分子子结构特征. 首先, 回顾 Transformer 中的注意力机制: 通过查询对象 (\mathbf{Q}) 从被查询对象 (\mathbf{V}) 中筛选出重要信息, 即计算查询对象和被查询对象中每个信息的相关程度. 其公式为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V}, \quad (3)$$

其中 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别代表查询向量、键向量和值向量, d_k 是向量的维度.

本文中采用随机初始化的 M 个可学习的特征作为查询向量 $\mathbf{y}_Q \in \mathbb{R}^{M \times \text{dim}}$, 并将经过分子表示学习模块后的原子节点特征 \mathbf{x} 作为键向量和值向量 $\mathbf{y}_K = \mathbf{y}_V = [\mathbf{x}_1^K, \mathbf{x}_2^K, \dots, \mathbf{x}_n^K]$. 所谓使用注意力机制, 即通过查询向量从原子节点特征中筛选出重要信息作为子结构. 式 (3) 在本文的场景中被写为

$$\mathbf{V} = \text{Softmax} \left(\frac{\mathbf{y}_Q \mathbf{W}_Q (\mathbf{y}_K \mathbf{W}_K)^T}{\sqrt{d}} \right) (\mathbf{y}_V \mathbf{W}_V), \quad (4)$$

其中 $\mathbf{W}_Q, \mathbf{W}_K$ 和 \mathbf{W}_V 分别表示查询、键和值的可学习权重矩阵. 然后将得到的注意力结果和查询向量相加, 并经过一层可学习参数为 \mathbf{W}_O 的前馈神经网络得到最后的子结构矩阵:

$$\mathbf{O} = \text{ReLU}((\mathbf{y}_Q \mathbf{W}_Q + \mathbf{V}) \mathbf{W}_O). \quad (5)$$

最终提取的子结构形式为矩阵 $\mathbf{O} = [o_1, o_2, \dots, o_M] \in \mathbb{R}^{M \times \text{dim}}$, 由 M 个维度为 dim 的向量组成, 其中每个向量与一个分子子结构模式相关联. 对于特定的药物对 (d_1, d_2) , 本文使用上述过程识别它们各自的 M 个分子子结构特征向量. 随后, 评估两种药物之间每对代表性向量的余弦相似度, 从而得到一个相似性矩阵 $\mathbf{S} \in \mathbb{R}^{M \times M}$. 具体而言, \mathbf{S} 中的每个元素计算如下:

$$\mathbf{S}_{ij} = \frac{(\mathbf{o}_i^1)^T \mathbf{o}_j^2}{\|\mathbf{o}_i^1\|_2 \|\mathbf{o}_j^2\|_2}, \quad (6)$$

其中, \mathbf{o}_i^1 (对应 \mathbf{o}_j^2) 表示第 i 个 (对应第 j 个) 代表性分子子结构特征向量. \mathbf{S} 中元素的值可以反映两种药物的一对分子子结构之间相互作用的强度, 从而增强了模型的解释性. \mathbf{S} 将在后文介绍的预测模块中使用, 用于进行最终的预测.

DDI 网络学习模块. 研究药物相互作用网络对于预测特定药物对的相互作用具有重要意义, 因为它能够通过捕捉药物之间的全局关系和交互模式, 基于已有的药物对关系推断新的相互作用, 从而提高预测的准确性和鲁棒性. 通过前文介绍的分子表示学习模块和子结构相互作用模块, 模型现在可以获得其分子全结构级和分子子结构级的表示 \mathbf{x}_d 和 \mathbf{O}_d . 回顾在 3.1 小节介绍的 DDI 网络, 每个药物被视为 DDI 网络中的节点, 它们的相互作用关系由 \mathbf{A} 给出. 这里将分子子结构表示矩阵 \mathbf{O}_d 展平, 并将其与分子全结构级表示 \mathbf{x}_d 拼接起来, 作为 DDI 网络中药物初始节点特征:

$$\mathbf{h}_d^0 = \text{Concat}(\mathbf{x}_d, \text{Flatten}(\mathbf{O}_d)). \quad (7)$$

然后, 此处模型使用图神经网络 (GNN) 来学习 DDI 网络中的节点特征. 若将第 l 层 GNN 输出的节点特征矩阵表示为 \mathbf{H}^l , 节点特征矩阵的更新可以形式化写为

$$\mathbf{H}^l = \text{GNN}(\mathbf{H}^{l-1}, \mathbf{A}). \quad (8)$$

在本文实验中, GNN 选择了能处理稀疏连接大型数据的图注意力网络 (graph attention network, GAT) [55]. 其具体公式为

$$h'_i = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W} h_j \right), \quad (9)$$

其中 \mathcal{N}_i 表示节点 i 的邻居节点集合, h_j 是节点 j 的特征向量, \mathbf{W} 是共享的线性变换权重矩阵, α_{ij} 是注意力系数, 表示节点 j 对节点 i 的重要性程度, 其计算方式为 $e_{ij} = a(\mathbf{W} h_i, \mathbf{W} h_j)$, $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}$ 其中 a 是一个注意力机制函数, 例如可以是一个单层神经网络. 然而其他 GNN, 例如图卷积网络 (graph convolutional network, GCN) [56] 和图同构网络 (graph isomorphism network, GIN) [57], 也可以被选用.

预测模块. 当获得了多个层次的信息后, 本模块的目的即是最大化地利用它们来预测最后的结果. 对于任意需要预测的药物对 d_1 和 d_2 , 本模块负责预测它们之间可能发生类型为 t 的相互作用事件的概率. HierDDI 将在进行 DDI 网络学习模块之前和之后分别进行两次预测. 给定 (d_1, d_2) , 在经过 DDI 网络学习模块之前, 模型已经获得了药物 d_1 和 d_2 的分子全结构级表示 $\mathbf{x}_{d_1}, \mathbf{x}_{d_2}$, 以及从分子子结构表示矩阵 $\mathbf{O}_{d_1}, \mathbf{O}_{d_2}$ 计算得到的 d_1 和 d_2 相互作用强度矩阵 \mathbf{S} . 然后, 展平相互作用强度矩阵 \mathbf{S} , 将其与两种药物的分子全结构级表示和相互作用类型的 one-hot 编码 \mathbf{t} 拼接起来. 然后将拼接得到的向量输入进一个多层感知机 (multi-layer perceptron, MLP), 并通过 sigmoid 激活函数 $\sigma(\cdot)$ 得到 Score_1, 即

$$\text{Score}_1 = \sigma(\text{MLP}(\text{Concat}(\mathbf{x}_{d_1}, \mathbf{x}_{d_2}, \text{Flatten}(\mathbf{S})), \mathbf{t})). \quad (10)$$

第二次预测发生在通过 DDI 网络学习模块之后. 此时, 经过 DDI 网络的传播与迭代, 模型得到了另一组分子全结构级表示 (\mathbf{x}'_{d_1} 和 \mathbf{x}'_{d_2}) 以及分子子结构表示矩阵 (\mathbf{O}'_{d_1} 和 \mathbf{O}'_{d_2}), 它们源自 DDI 网络中的节点特征 \mathbf{h}'_{d_1} 和 \mathbf{h}'_{d_2} (按照维数拆开). 这里使用 \mathbf{O}'_{d_1} 和 \mathbf{O}'_{d_2} 并遵循式 (6), 可以计算一个新的相互作用强度矩阵 \mathbf{S}' . 与计算 Score_1 类似, 可以通过以下方式获得 Score_2:

$$\text{Score}_2 = \sigma(\text{MLP}(\text{Concat}(\mathbf{x}'_{d_1}, \mathbf{x}'_{d_2}, \text{Flatten}(\mathbf{S}')), \mathbf{t})). \quad (11)$$

模型将从两次预测中获得的概率值取平均作为最终的预测结果 \hat{y} :

$$\hat{y} = \frac{1}{2} (\text{Score}_1 + \text{Score}_2). \quad (12)$$

Score_1 和 Score_2 之间的区别在于用于预测 Score_1 的特征仅来自需要预测的药物对的结构信息, 而用于预测 Score_2 的特征则聚合了 DDI 网络中除了待预测的两种药物之外的其他药物的信息. 为了鼓励模型从待预测药物的分子全结构级和分子子结构级特征中更有效地提取药物自身结构信息, 此处特意保留了仅基于这两种药物本身信息进行预测得到的 Score_1.

3.2 训练过程

回顾 3.1 小节介绍的 DDI 预测, 本文将 DDI 预测任务视为一个二分类问题. 数据集 \mathcal{M} 只包含已知存在的 DDIs, 这里将其视为正样本. 然而, 为了有效地进行训练, 此处需要进行负采样以平衡正负样本比例. 本文采用了文献 [58] 中提出的方法: 给定一个已知的 DDI 三元组样本 (d_i, d_j, t) , 通过替换 d_i 或 d_j 生成一个负样本. 通过这种方法, 可以确保对于每种作用类型 t , 正样本和负样本之间保持平

算法 1 HierDDI 的训练算法.

输入: DDI 数据集 \mathcal{M} , 药物集合 \mathcal{D} , 超参数 K, L ;

输出: 模型 $\Gamma_\theta(d_i, d_j, t | \mathcal{M}, \mathbf{A}_N)$;

```

1: 从数据集  $\mathcal{M}$  中构建  $\mathbf{A}_N$ ;
2: 将每个药物表示为对应的分子图  $G = (\mathbf{X}, \mathbf{E})$ ;
3: while HierDDI 未收敛 do
4:   for  $\mathcal{M}$  中的每个三元组样本  $(d_1, d_2, t)$  do
5:     for  $\mathcal{D}$  中的每个药物  $d$  do
6:       初始化药物  $d$  的节点特征矩阵  $\mathbf{X}_d^0$  和边特征矩阵  $\mathbf{E}$ ;
7:       参考式 (1) 和 (2) 进行  $K$  次消息传递;
8:       获得每个原子节点特征  $\{x_1^K, x_2^K, \dots, x_n^K\}$ ;
9:       计算分子全结构级特征  $\mathbf{x}_d = \frac{1}{n} \sum_p (\mathbf{x}_p^K)$ ;
10:      参考式 (3)~(5) 计算分子子结构表示矩阵  $\mathbf{O}_d$ ;
11:      参考式 (7) 得到药物特征  $\mathbf{h}_d^0$ ;
12:    end for
13:    参考式 (6) 和 (10) 计算得到  $(d_1, d_2, t)$  样本的 Score.1;
14:    将所有药物的特征  $\{\mathbf{h}_d^0 | d \in \mathcal{D}\}$  用于初始化  $\mathbf{H}^0$ ;
15:    参考式 (9) 进行  $L$  层图神经网络传播;
16:    获得所有药物更新后的特征  $\{\mathbf{h}'_d | d \in \mathcal{D}\}$ ;
17:    将更新后  $d_1, d_2$  的药物特征  $\mathbf{h}'_{d_1}, \mathbf{h}'_{d_2}$  拆解成分子全结构级  $\mathbf{x}'_{d_1}, \mathbf{x}'_{d_2}$  和分子子结构级  $\mathbf{O}'_{d_1}, \mathbf{O}'_{d_2}$ ;
18:    参考式 (6) 和 (11) 计算得到  $(d_1, d_2, t)$  样本的 Score.2;
19:    计算预测概率  $y = (\text{Score.1} + \text{Score.2})/2$ ;
20:    参考式 (13) 计算损失函数;
21:    计算梯度, 并进行反向传播更新模型参数  $\theta$ ;
22:  end for
23: end while
24: 返回  $\Gamma_\theta$ .
```

衡. 为了优化模型的可学习参数, 本文使用二元交叉熵 (binary cross entropy, BCE) 损失函数:

$$L = -\frac{1}{2N} \sum_{i=1}^{2N} (y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)), \quad (13)$$

其中, $y_i \in \{0, 1\}$ 是第 i 组样本的真实标签, \hat{y}_i 是第 i 组样本相互作用的预测概率, N 是已知 (正样本) 药物相互作用关系的数量. 有关 HierDDI 的训练过程参照算法 1.

3.3 复杂度分析

假设数据集中的药物数量为 N , 药物相互作用的数量为 I , 药物分子中包含的平均原子数量为 A , 药物分子中包含的平均化学键数量为 B , 原子特征维度为 d . 对于分子表示学习模块, 获取分子全结构级表示的时间复杂度为 $\mathcal{O}(KBd^2)$, 其中 K 为消息传播的次数. 对于子结构相互作用模块, 提取分子子结构级表示的时间复杂度为 $\mathcal{O}(M^2d)$, 其中 M 是子结构模式个数. 由于本文可以同时处理 N 个药物并行处理, 所以这两个模块处理 N 个分子的时间复杂度与处理一个分子是相同的. 对于 DDI 网络学习模块, 其时间复杂度与药物相互作用数量成正比, 为 $\mathcal{O}(LI d^2)$. 预测模块的时间复杂度为 $\mathcal{O}(Id)$. 整个算法的复杂度为 $\mathcal{O}((KB + LI)d^2 + (M^2 + I)d)$. 从公式可见, 算法的复杂度与数据集中药物相互作用的数量呈正相关, 同时, 分子中化学键的数量、消息传播的次数、图神经网络的层数以及子结构模式个数也会影响算法的复杂度.

4 实验与讨论

本节在两个场景下通过与现有方法进行对比, 全面评估了 HierDDI 的性能. 所有实验都在一台配备了 i9-10920X CPU, RTX 3090 GPU 和 128G RAM 的微型服务器上运行.

表 2 本文所使用数据集的统计信息.
Table 2 Statistics of the used datasets.

Item	DrugBank	Twosides
#Drugs	1706	645
#Interactions	191808	5676287
#Relation types	86	963

4.1 实验设置

数据集. 本文使用了两个真实的数据集来评估所提出的方法. (1) DrugBank^[59] 是来自 FDA/Health Canada 药品标签的数据. DrugBank 中的不同相互作用类型描述了一种药物如何影响另一种药物的代谢; (2) Twosides^[60] 是在过滤原始 Twosides 副作用数据后提出的. 与 DrugBank 数据集不同, Twosides 中的相互作用类型更多的是在表型水平而不是代谢水平上, 这意味着相互作用只是副作用. 与现有工作^[30] 的做法保持一致, 为了只保留出现频率较高的作用类型, 此处过滤掉了那些在 Twosides 数据集中出现次数不到 500 的作用类型. 表 2 显示了它们的统计信息.

指标与基线模型. 与之前的工作^[27,29] 一致, 本文采用以下指标进行评价: 准确率 (accuracy, ACC)、受试者操作曲线下面积 (area under curve, AUC)、平均精度 (average precision, AP)、F1 得分、精确率 (Precision) 和召回率 (Recall). 本文将提出的 HierDDI 与现有领先方法进行比较. 这里选择了 6 个基线模型: GAT-DDI^[55] 直接应用图注意力网络 GAT 进行分子表示学习, 并将得到的特征直接用于预测是否发生反应作用. MUFFIN^[27] 结合了药物自身结构信息和具有丰富生物学信息的知识图谱 (KG) 来预测最终结果. GMPNN^[30] 引入了门控消息传递网络, 它能学习不同大小和形状的分子子结构. SSI-DDI^[21] 将每个节点的隐藏特征视为子结构, 然后计算这些子结构之间的相互作用来进行最终的 DDI 预测. MSAN^[29] 提出了一个子结构注意力感知模块, 用于学习分子图的固定数量的分子子结构模式. DSN-DDI^[31] 提出了一种双视角学习模型, 从单个药物 (“内视图”) 和药物对 (“间视图”) 中学习药物子结构. 这里 GAT-DDI 基线是由作者根据图注意力网络进行复现的, 其对应参数与训练策略与本文提出的 HierDDI 相同. 至于 MUFFIN, GMPNN 和 SSI-DDI, 这 3 个基线模型的原作者提供了对应的代码, 本文直接使用其开源代码进行重训练并在本文使用的数据集上进行测试, 其对应参数选择也依照原作者的描述. 而 MSAN 和 DSN-DDI 基线模型的原作者均提供了对应的代码和训练好的模型, 所以本文直接使用其提供的模型进行测试.

实验场景. 本文考虑了两种不同的场景, 分别为直推式场景和归纳式场景.

直推式 (transductive) 场景. 在直推式场景中, 训练集中的药物也会出现在测试集中. 本文从 DDI 三元组样本的角度对两个数据集按照训练集 60%、验证集 20%、测试集 20% 的比例进行划分, 并保持各个集合中相互作用类型的比例一致. 在这种场景下, 利用训练集中的 DDIs 构建 DDI 网络矩阵, 该矩阵将用于训练和推断阶段.

归纳式 (inductive) 场景. 与直推式场景不同的是, 这里的数据集是基于药物进行拆分的, 这意味着训练集和测试集的 DDI 样本中不具有重叠的药物. 这在文献中通常被称为冷启动场景, 且与直推式场景相比起来更难处理. 参考文献^[21] 的做法, 首先将所有药物按 1:4 的比例拆分为 \mathcal{G}_{new} 和 \mathcal{G}_{old} , 然后训练集仅包含两种药物都在 \mathcal{G}_{old} 中的 DDI 样本. 测试集 \mathcal{M}_{S_1} 包含两种药物都在 \mathcal{G}_{new} 中的 DDI 样本, 而测试集 \mathcal{M}_{S_2} 则包含两种药物中有且仅有一种在 \mathcal{G}_{new} 中的 DDI 样本. 需要注意的是, 由于训练集和测试集的 DDI 样本中没有重叠的药物, 所以如果依旧使用训练集中的 DDIs 来构建 DDI 网络矩阵, 那么 \mathcal{G}_{new} 中的药物将在 DDI 网络拓扑结构中全部作为孤点存在, DDI 网络学习模块将失效于这种情况. 所以这里除了使用训练集中的 DDI 来构建 DDI 网络矩阵外, 还额外在新药物和已见过的药物之间建立随机边, 并添加进 DDI 网络矩阵. 这确保了 DDI 网络学习模块可以推广到在训练集中

表 3 在直推式场景下,各方法在 DrugBank 和 Twosides 上的性能 (平均值 \pm 标准差). 每个指标中最佳性能以粗体字显示. 当某一方法的准确率 (ACC) 低于 50% 时,该方法的所有指标都被标记为 “-”.

Table 3 Performance (mean \pm std) on DrugBank and Twosides in the transductive setting. The best performance in each metric is in bold. If the accuracy (ACC) of a method is below 50%, all metrics are marked as “-”.

Model	DrugBank					Twosides				
	ACC	AUC	F1	Precision	Recall	ACC	AUC	F1	Precision	Recall
GAT-DDI	89.91 \pm 1.00	95.21 \pm 0.70	90.20 \pm 0.96	87.04 \pm 1.11	93.56 \pm 0.52	-	-	-	-	-
MUFFIN	93.90 \pm 0.08	98.29 \pm 0.13	93.99 \pm 0.07	93.06 \pm 0.12	94.94 \pm 0.14	82.79 \pm 0.09	90.50 \pm 0.10	82.85 \pm 0.09	82.54 \pm 0.08	83.17 \pm 0.11
GMPNN	95.30 \pm 0.05	98.46 \pm 0.01	95.38 \pm 0.09	93.60 \pm 0.07	97.22 \pm 0.10	82.83 \pm 0.14	90.07 \pm 0.12	84.02 \pm 0.05	78.42 \pm 0.11	90.61 \pm 0.23
SSI-DDI	96.33 \pm 0.09	98.95 \pm 0.08	96.38 \pm 0.09	95.09 \pm 0.08	97.70 \pm 0.14	78.20 \pm 0.14	85.85 \pm 0.13	79.05 \pm 0.20	74.33 \pm 0.21	86.15 \pm 0.15
MSAN	97.00 \pm 0.09	99.27 \pm 0.03	97.04 \pm 0.08	95.92 \pm 0.19	98.18\pm0.09	83.73 \pm 0.04	91.26 \pm 0.06	85.94 \pm 0.08	79.90 \pm 0.07	92.27 \pm 0.11
DSN-DDI	96.91 \pm 0.03	99.46 \pm 0.02	96.82 \pm 0.03	96.22 \pm 0.03	97.43 \pm 0.03	83.21 \pm 0.02	90.93 \pm 0.02	85.02 \pm 0.02	79.12 \pm 0.03	91.20 \pm 0.12
HierDDI	97.58\pm0.12	99.66\pm0.07	97.58\pm0.09	97.67\pm0.13	97.49 \pm 0.14	85.75\pm0.12	92.45\pm0.07	86.72\pm0.12	80.94\pm0.09	93.39\pm0.10

表 4 在归纳式场景下,各方法在 DrugBank 和 Twosides 上的性能 (平均值 \pm 标准差). 每个指标中最佳性能以粗体字显示. 当某一方法的准确率 (ACC) 低于 50% 时,该方法的所有指标都被标记为 “-”.

Table 4 Performance (mean \pm std) on DrugBank and Twosides in the inductive setting. The best performance in each metric is in bold. If the accuracy (ACC) of a method is below 50%, all its metrics are marked as “-”.

	DrugBank				Twosides				
	ACC	AUC	F1	AP	ACC	AUC	F1	AP	
\mathcal{M}_{S1} partition	GAT-DDI	66.31 \pm 0.61	72.75 \pm 0.78	68.68 \pm 0.60	71.61 \pm 1.00	-	-	-	-
	MUFFIN	63.57 \pm 1.12	69.60 \pm 1.58	60.46 \pm 1.30	67.12 \pm 1.54	-	-	-	-
	GMPNN	68.57 \pm 0.30	74.96 \pm 0.40	65.32 \pm 0.23	75.44 \pm 0.50	53.20 \pm 1.21	53.98 \pm 1.23	30.12 \pm 5.21	51.32 \pm 1.34
	SSI-DDI	65.40 \pm 1.30	73.43 \pm 1.81	54.12 \pm 3.46	75.03 \pm 1.42	50.84 \pm 0.34	52.77 \pm 1.04	22.05 \pm 3.21	51.92 \pm 0.23
	MSAN	69.23 \pm 0.93	76.23 \pm 0.79	67.20 \pm 1.98	77.07 \pm 0.89	52.01 \pm 1.04	52.65 \pm 1.90	21.74 \pm 6.21	51.88 \pm 1.20
	DSN-DDI	73.42 \pm 1.29	81.79 \pm 1.12	70.34 \pm 0.98	81.82 \pm 1.48	53.21 \pm 1.24	54.77 \pm 1.24	30.94 \pm 7.21	51.38 \pm 2.20
	HierDDI	77.62\pm2.23	84.20\pm2.12	76.99\pm1.03	84.13\pm1.34	56.73\pm4.21	56.85\pm3.02	37.35\pm8.31	52.71\pm3.20
\mathcal{M}_{S2} partition	GAT-DDI	69.83 \pm 1.41	77.29 \pm 1.63	73.01 \pm 0.85	75.79 \pm 1.95	-	-	-	-
	MUFFIN	75.20 \pm 1.22	84.35 \pm 0.89	72.15 \pm 1.12	83.63 \pm 1.54	56.43 \pm 0.91	62.91 \pm 1.02	40.42 \pm 1.33	60.16 \pm 0.89
	GMPNN	77.72 \pm 0.30	84.84 \pm 0.15	78.29 \pm 0.16	84.87 \pm 0.40	57.04 \pm 1.23	64.21\pm2.10	48.12 \pm 3.01	61.12 \pm 0.79
	SSI-DDI	76.38 \pm 0.92	84.23 \pm 1.05	73.54 \pm 1.50	84.94 \pm 0.76	57.63 \pm 0.82	63.54 \pm 2.02	47.61 \pm 4.27	60.63 \pm 0.82
	MSAN	77.61 \pm 0.45	85.69 \pm 0.23	76.65 \pm 1.03	85.36 \pm 0.67	57.27 \pm 0.93	63.38 \pm 2.30	44.13 \pm 2.88	61.46\pm0.91
	DSN-DDI	81.92 \pm 1.20	91.01\pm0.76	80.18 \pm 1.49	91.09\pm0.93	57.03 \pm 0.93	62.13 \pm 1.83	46.23 \pm 3.28	61.23 \pm 1.11
	HierDDI	82.55\pm1.06	89.90 \pm 2.03	81.66\pm0.37	87.30 \pm 1.22	60.92\pm1.39	59.36 \pm 1.92	55.73\pm6.81	58.85 \pm 1.07

没有见过的药物.

对于每个 DDI 样本,可以根据 3.2 小节中讨论的方法生成对应的负样本. 这里负样本均在训练之前生成,以确保所有模型都在相同的数据上训练. 在两种场景中,重复数据划分 3 次,并针对每个数据划分进行了 5 次独立实验. 本文报告了这些实验中每个模型对应结果的平均值和标准差. 在直推式场景中,本文使用 Adam 优化器训练模型 300 轮,初始学习率设置为 $1e-4$,在经过 200 轮后衰减为原始值的 $1/10$. 在归纳式场景下,由于测试集和训练集之间的分布不同,如果训练过多轮可能会导致过拟合,从而在测试集上表现下降. 因此,在归纳式场景中总共只训练 100 轮. 在 DrugBank 和 Twosides 数据集上,批量大小分别设置为 256 和 512.

4.2 实验结果与分析

这一部分比较了 HierDDI 与基线模型的性能. 表 3 和 4 中分别展示了在直推式和归纳式场景下的实验结果. 由于数据集 Twosides 规模更大,且在同一对药物之间存在多种相互作用关系 (通常被称为多标签任务),所以相比 DrugBank (单标签多分类任务数据集),所有模型的表现均有所下降. 而实

表 5 在 DrugBank 数据集上进行的消融实验. 每个指标中最佳性能以粗体字表示.

Table 5 Ablation studies on DrugBank dataset. The best performance in each metric is in bold.

Approach	Transductive			Inductive \mathcal{M}_{S1}			Inductive \mathcal{M}_{S2}		
	ACC	AUPR	F1	ACC	AUPR	F1	ACC	AUPR	F1
HierDDI (our full version)	85.75±0.12	92.45±0.07	86.72±0.12	56.73±4.21	56.85±3.02	37.35±8.31	60.92±1.39	59.36±1.92	55.73±6.82
- substructure	96.87±0.32	99.01±0.21	96.89±0.33	72.18±4.25	80.39±3.20	76.02±1.29	78.22±3.25	87.14±2.63	79.66±1.02
- DDI network	96.25±0.07	98.22±0.10	95.21±0.09	68.13±0.63	76.02±0.68	67.01±0.87	77.13±0.35	85.09±0.23	75.85±0.43
- substructure & DDI network	91.62±0.06	95.82±0.08	90.20±0.06	66.71±0.35	72.85±0.29	69.08±0.30	70.43±0.61	78.43±1.22	73.91±0.35

验结果表明, HierDDI 在两个数据集的直推式和归纳式场景中均达到了最优性能. 具体而言, 在直推式场景中, 相较于领先方法, 本文提出的方法在数据集 DrugBank 和 Twosides 的预测准确率上分别提升了 0.58% 和 2.02%. 在更具挑战性的归纳式场景下, HierDDI 的准确率提升更为显著, 高达 4.20% 和 3.52%.

值得注意的是, 基线模型 MSAN, SSI-DDI 以及 GMPNN 也提出了学习分子子结构的方法, 甚至在对子结构的处理上引入了更复杂的模块. 本文提出的模型表现优于这些方法, 这表明通过 DDI 网络引入更宏观的视角对整体预测非常有效. 另外, 基线模型 MUFFIN 也整合了多层级特征, 但所有这些特征都是事先使用独立网络训练提取的. 从结果可见, HierDDI 的性能明显优于 MUFFIN, 表明使用端到端的训练范式能放大 3 个层次作为一个整体的集体效用, 体现了端到端训练模型相较于独立网络分步训练模型更为优越.

此外, 作者发现所有方法在归纳式中都表现出明显的性能下降, 在 Twosides 数据集上部分模型甚至失效. 这很大程度上由于 Twosides 数据集包含的总药物少且训练集和测试集中的 DDI 样本没有重叠的药物. 具体而言, 即使使用文献 [61] 中提出的方法, 在 \mathcal{G}_{new} 中找到每个药物与 \mathcal{G}_{old} 中最近邻的药物, 它们之间的平均子结构相似度仅约为 0.46. 在归纳式场景下, 尽管所有方法的性能均下降不少, 本文提出的模型却能大幅提升原有基线的预测水平, 可见提取多层次的建模方式能提升冷启动场景性能, 尤其是对比其他基线方法, 同时建模子结构级和 DDI 网络级的特征增强了模型的鲁棒性, 使得其能在归纳式场景中依旧表现良好. 值得注意的是, 为了利用现有药物相互作用网络结构来辅助新药物之间相互作用的预测, 本文在构建 DDI 网络时在新药物与现有已知药物之间添加了随机边, 这使得新药物节点不再是网络中的孤点, 也就是新药物的特征能通过网络的拓扑结构被现有药物多次更新. 此外, 笔者观察到, 本文提出的模型在归纳式场景下显示相对较大的方差, 作者认为上述添加的随机边带来的噪声是造成不稳定性一个可能原因.

4.3 消融实验

本文进一步探讨了提出的 HierDDI 模型各组件对最终性能的影响. 本文在两个数据集上均进行了消融实验, 结果如表 5 和 6 所示. 删去子结构相互作用模块的模型 (-substructure) 在计算 Score_1 和 Score_2 时仅使用分子全结构级别特征, 而删去 DDI 网络学习模块的模型 (-DDI network) 则仅使用 Score_1 作为最终预测结果. 同时删去这两个模块的模型 (-substructure & DDI network) 只包含消息传递神经网络来学习分子全结构的特征, 并直接用于预测得到最终结果. 由于同时删除这两个模块会使模型在 Twosides 数据集上的准确率低于 50%, 即实验失败, 因此本文未报告相应实验结果. 从表中结果可见, 所有模块, 尤其是 DDI 网络学习模块, 对模型性能均有积极贡献. 去除子结构相互作用模块会增加性能波动, 凸显了分子子结构特征在提升模型稳定性中的重要性.

4.4 进一步验证

子结构模式个数 M 的影响. 为了进一步探索在 HierDDI 中子结构模式数量的影响, 此处改变了子结构模式的数量 M 进行实验. 值得注意的是, 这里的实验删去了 DDI 网络学习模块, 以便对子结构相互作用模块中的 M 这一变量进行更为清晰的比较. 结果如图 3(a) 所示. 作者观察到当 M 的大

表 6 在 Twosides 数据集上进行的消融实验. 每个指标中最佳性能以粗体字表示.

Table 6 Ablation studies on Twosides dataset. The best performance in each metric is in bold.

Approach	Transductive			Inductive \mathcal{M}_{S1}			Inductive \mathcal{M}_{S2}		
	ACC	AUPR	F1	ACC	AUPR	F1	ACC	AUPR	F1
HierDDI (our full version)	97.49±0.12	99.61±0.07	97.48±0.09	77.62±2.23	84.20±2.12	76.99±1.03	82.55±1.06	89.90±2.03	81.66±0.37
- substructure	83.52±0.42	91.09±0.29	84.91±0.41	53.78±3.16	55.90±3.56	31.24±3.83	56.37±0.32	60.34±1.97	48.83±2.76
- DDI network	83.01±0.39	90.23±0.13	84.40±0.28	53.12±0.56	54.32±1.39	28.39±1.49	56.97±0.23	59.10±1.02	46.23±1.54

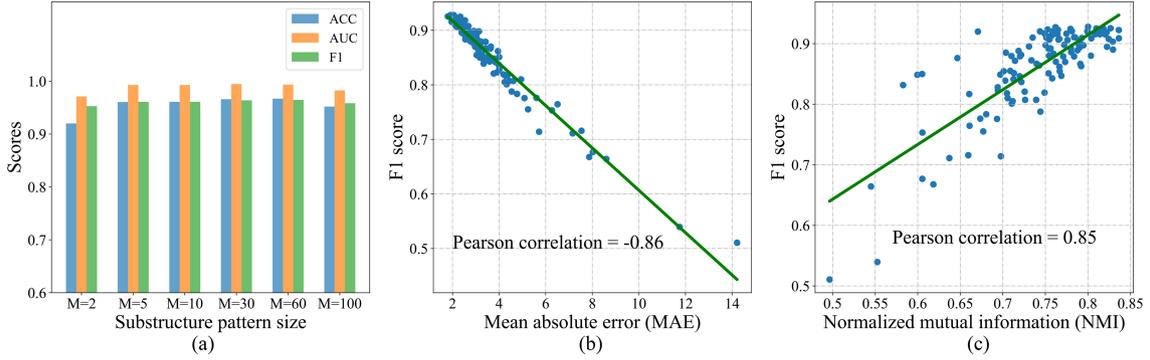


图 3 (网络版彩图) (a) 在不同子结构数量 M 下的结果; (b) 模型性能 (F1 得分) 与网络度恢复的 MAE 之间的 Pearson 相关系数; (c) 模型性能 (F1 得分) 与网络社区恢复的 NMI 之间的 Pearson 相关系数.

Figure 3 (Color online) (a) Results with varying sizes of substructure pattern M ; (b) Pearson correlation between prediction (F1) and MAE of degree recovery; (c) Pearson correlation between prediction (F1) and NMI of community recovery.

小范围从 10 到 60 时, 模型的性能相对稳定, 都有较好的效果. 若 M 太小, 子结构相互作用模块几乎不起作用, 而当 M 持续增大时, 模型效果并不会继续提升, 反而会使得特征维数增加导致在计算上变得更加耗时.

网络结构预测中的度和社区恢复. 最近几项研究^[62~64] 揭露了考虑网络中的度 (degree) 和社区 (community) 结构等特性在网络链接预测 (link prediction) 方面的有效性. 从 DDI 网络的角度来看, 预测药物相互作用是否存在其实就是 DDI 网络上的链接预测. 从这个角度, 本文提出的 HierDDI 本质上是利用网络拓扑结构来有效地发现网络中缺失的链接, 从而恢复 DDI 网络的度和社区划分. 图 3(b) 和 (c) 展示了度和社区的恢复程度对链接预测的影响. 这里随机选择了不同训练阶段的模型进行测试, 得到其对应的链路预测结果, 并计算了预测度和真实度之间的平均绝对误差 (mean absolute error, MAE) 来表示度的恢复程度. 至于社区恢复, 首先使用 Louvain 社区检测算法^[65] 将药物标记为不同的社区, 然后使用归一化互信息 (normalized mutual information, NMI) 来衡量社区的恢复程度. Louvain 算法是一种高效的社区发现算法, 通过优化模块度来检测网络中的社区结构. 其基本思想是网络中节点尝试遍历所有邻居的社区标签, 并选择最大化模块度增量的社区标签. 在最大化模块度之后, 把每个社区看成一个新的节点, 重复直到模块度不再增大. 归一化互信息是一种常见的聚类有效性外部评价指标, 从信息论的角度评估了两个聚类结果的相似性. 设聚类的实验结果为 X , 真实结果为 Y , 则其计算式如下:

$$\text{NMI}(X, Y) = \frac{I(X, Y)}{\sqrt{H(X)H(Y)}}, \quad (14)$$

其中, $I(X, Y)$ 表示 X 和 Y 之间的互信息, $H(X)$ 和 $H(Y)$ 表示 X 和 Y 的熵. NMI 的取值范围为 $[0, 1]$, 值越大表明和真实结果的共享信息越多, 即聚类效果越好. 皮尔逊相关系数 (Pearson correlation coefficient) 是用于衡量两个变量之间线性相关程度的统计指标, 并且可以用于连续型数据. 它的取值范围是 $[-1, 1]$, 其中当值为 1 时, 表示两个变量完全线性正相关, 当值为 -1 时则表示两个变量完全线性负相关. 这里本文选用 F1 值来代表模型表现. 从图中可见, 度恢复和模型性能之间的 Pearson 相关

表 7 本文模型所验证的潜在 DDIs. 这些 DDIs 未包含在本文使用的数据集中, 但已经验证并在官方数据库中记录.

Table 7 Potential DDIs verified by our model. They are not in the used dataset, yet verified and recorded in official databases.

Drug 1	Drug 2	Description	Probability
Bivalirudin	Carmustine	Increase the risk or severity of bleeding	0.9806
Goserelin	Glisoexepide	Decrease the therapeutic efficacy	0.9999
Timolol	Goserelin	Increased the risk or severity of QTc prolongation	0.9950
Donepezil	Nifedipine	Increase the bradycardic activities	0.9975
Sulfamoxole	Donepezil	Decrease the metabolism	0.9993
Roxatidine acetate	Cysteamine	Decrease the bioavailability	0.9935
Lofexidine	Perphenazine	Increase the therapeutic efficacy	0.9999
Dacarbazine	Sunitinib	Increase the risk or severity of adverse effects	0.9786
Levosaltamol	Pseudoephedrine	Increase the risk or severity of hypertension	0.9828
Pseudoephedrine	Methantheline	Decrease the sedative activities	0.9995

系数为 -0.86 , 社区恢复和模型性能之间的值为 0.85 , 这也进一步验证了更好的网络结构恢复能力意味着更好的 DDI 预测性能.

预测潜在的 DDIs. DDIs 数据集已知存在不完整性, 许多潜在的 DDIs 目前尚未被记录. 为展示本文提出的模型面对现实场景的能力, 表 7 中展示了使用 DrugBank 训练的模型预测的 10 条药物相互作用. 这些相互作用未包含在本文使用的 DrugBank 数据集中, 但它们已经经过验证, 并可在官方 DrugBank 数据库 (<https://go.DrugBank.com>) 中查阅.

5 结论与展望

药物相互作用可以带来积极正向的治疗结果, 也可能会引起严重的不良反应. 近年来药物规模大幅增加, 对药物之间相互作用类型的研究需求也日益剧增. 在这样的背景下, 构建机器学习模型进行自动化预测反应类型变得尤为重要. 然而此前基于机器学习的方法均未能全面地从多层次建模药物相互作用问题, 且部分方法还需依赖生物医学知识图谱. 为此, 本文提出了 HierDDI, 一种新颖的方法旨在通过多层次结构建模以有效地学习药物间的相互作用, 且即使在没有生物医学知识图谱的情况下也能够有效预测. 本文提出的方法跨越 3 个不同的层次提取和整合信息: 分子子结构、分子全结构和 DDI 网络结构. 这种多层次建模是通过端到端的训练模型实现的, 每个层次相互增强, 从而全面理解 DDIs. 广泛的实验表明本文提出的方法在性能上明显优于现有领先方法, 成为目前表现最佳的模型. 在直推式场景下, 在 DrugBank 和 Twosides 数据集上的预测准确率相较于领先方法分别提升了 0.58% 和 2.02% , 在归纳式场景下提升则达到了 4.20% 和 3.52% .

未来研究拟将 HierDDI 拓展至药物与靶标、药物与疾病、基因与疾病关联等领域. 借助类似多层次建模方法, 深入剖析基因、疾病、药物和靶标间的相互作用, 为个性化治疗与药物设计提供思路. 此外, 还将探索 HierDDI 与临床实践的融合路径, 通过与医疗机构、药品监管部门合作, 验证模型准确性与安全性, 将其整合至临床决策支持系统, 以提升患者治疗效果与生活质量.

参考文献

- 1 Hoogeboom E, Satorras V G, Vignac C, et al. Equivariant diffusion for molecule generation in 3D. In: Proceedings of International Conference on Machine Learning, 2022. 8867–8887
- 2 Yang N, Wu H, Zeng K, et al. Molecule generation for drug design: a graph learning perspective. Fundamental Res, 2024. doi: 10.1016/j.fmre.2024.11.027

- 3 Wu H, Ye X, Yan J. QVAE-mole: the quantum VAE with spherical latent variable learning for 3-D molecule generation. In: Proceedings of the 38th Annual Conference on Neural Information Processing Systems, 2024
- 4 Dauparas J, Anishchenko I, Bennett N, et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 2022, 378: 49–56
- 5 Watson J L, Juergens D, Bennett N R, et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 2023, 620: 1089–1100
- 6 Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 2024, 630: 493–500
- 7 Wu H, Liu W, Bian Y, et al. EBMdock: neural probabilistic protein-protein docking via a differentiable energy model. In: Proceedings of the 12th International Conference on Learning Representations, 2024
- 8 Igashov I, Schneuing A, Segler M, et al. RetroBridge: modeling retrosynthesis with Markov bridges. In: Proceedings of the 12th International Conference on Learning Representations, 2024
- 9 Zeng K, Yang B, Zhao X, et al. Ualign: pushing the limit of template-free retrosynthesis prediction with unsupervised SMILES alignment. *J Cheminform*, 2024, 16: 80
- 10 Palleria C, Di Paolo A, Giofrè C, et al. Pharmacokinetic drug-drug interaction and their implication in clinical management. *J Res Med Sci*, 2013, 18: 601
- 11 Vilar S, Harpaz R, Uriarte E, et al. Drug-drug interaction through molecular structure similarity analysis. *J Am Med Inform Assoc*, 2012, 19: 1066–1074
- 12 Vilar S, Uriarte E, Santana L, et al. Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nat Protoc*, 2014, 9: 2147–2163
- 13 Liu Z J, Fu D X, Sun C H, et al. Advances of in vivo drug interaction studies. *Adv Drug React J*, 2006, 8: 33–38 [刘治军, 傅得兴, 孙春华, 等. 体内药物相互作用研究进展. *药物不良反应杂志*, 2006, 8: 33–38]
- 14 Tatonetti N P, Fernald G H, Altman R B. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *J Am Med Inform Assoc*, 2012, 19: 79–85
- 15 Lan X, Wei R, Cai H W, et al. Application of machine learning algorithm in medical field. *Chin Med Eq J*, 2019, 40: 93–97 [兰欣, 卫荣, 蔡宏伟, 等. 机器学习算法在医疗领域中的应用. *医疗卫生装备*, 2019, 40: 93–97]
- 16 Cao X M, Shi R G. Multi-task based neural network algorithm for detection of drug adverse event. *Control Eng China*, 2020, 27: 1151–1156 [曹晓民, 史瑞刚. 多任务神经网络药物不良反应检测算法. *控制工程*, 2020, 27: 1151–1156]
- 17 Yu L, Zhao J. Prediction of disease-drug relationships based on tissue specificity and direct neighbor similarity. *Sci Sin Inform*, 2019, 49: 1175–1185 [鱼亮, 赵晋. 基于组织特异性和直接邻居相似度方法预测疾病-药物关系. *中国科学: 信息科学*, 2019, 49: 1175–1185]
- 18 Shi J Y, Mao K T, Yu H, et al. Detecting drug communities and predicting comprehensive drug-drug interactions via balance regularized semi-nonnegative matrix factorization. *J Cheminform*, 2019, 11: 28
- 19 Yu H, Mao K T, Shi J Y, et al. Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC Syst Biol*, 2018, 12: 14
- 20 Lin X, Quan Z, Wang Z J, et al. KGNN: knowledge graph neural network for drug-drug interaction prediction. In: Proceedings of IJCAI, 2020. 2739–2745
- 21 Nyamabo A K, Yu H, Shi J Y. SSI-DDI: substructure-substructure interactions for drug-drug interaction prediction. *Brief Bioinf*, 2021, 22: bbab133
- 22 Li P, Huang C, Fu Y, et al. Large-scale exploration and analysis of drug combinations. *Bioinformatics*, 2015, 31: 2007–2016
- 23 Zhou Y, Wang F, Tang J, et al. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digital Health*, 2020, 2: 667–676
- 24 Bumgardner B, Tanvir F, Saifuddin K M, et al. Drug-drug interaction prediction: a purely smiles based approach. In: Proceedings of IEEE International Conference on Big Data, 2021. 5571–5579
- 25 Hung T N K, Le N Q K, Le N H, et al. An AI-based prediction model for drug-drug interactions in osteoporosis and Paget's diseases from SMILES. *Mol Inf*, 2022, 41: 2100264
- 26 Chen X, Liu X, Wu J. Drug-drug interaction prediction with graph representation learning. In: Proceedings of IEEE International conference on bioinformatics and biomedicine (BIBM), 2019. 354–361
- 27 Chen Y, Ma T, Yang X, et al. MUFFIN: multi-scale feature fusion for drug-drug interaction prediction. *Bioinformatics*, 2021, 37: 2651–2658
- 28 Yang N, Zeng K, Wu Q, et al. Learning substructure invariance for out-of-distribution molecular representations. In: Proceedings of Advances in Neural Information Processing Systems, 2022. 35: 12964–12978
- 29 Zhu X, Shen Y, Lu W. Molecular substructure-aware network for drug-drug interaction prediction. In: Proceedings of

- the 31st ACM International Conference on Information & Knowledge Management, 2022. 4757–4761
- 30 Nyamabo A K, Yu H, Liu Z, et al. Drug-drug interaction prediction with learnable size-adaptive molecular substructures. *Brief Bioinf*, 2022, 23: bbab441
 - 31 Li Z, Zhu S, Shao B, et al. DSN-DDI: an accurate and generalized framework for drug-drug interaction prediction by dual-view representation learning. *Brief Bioinf*, 2023, 24: bbac597
 - 32 Zeng A G X, Bansal S, Jin L, et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat Med*, 2022, 28: 1212–1223
 - 33 Couturier C P, Ayyadhury S, Le P U, et al. Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat Commun*, 2020, 11: 3406
 - 34 Wigbers M C, Tan T H, Brauns F, et al. A hierarchy of protein patterns robustly decodes cell shape information. *Nat Phys*, 2021, 17: 578–584
 - 35 Zhang Y, Qu P, Ji Y, et al. A system hierarchy for brain-inspired computing. *Nature*, 2020, 586: 378–384
 - 36 Siegle J H, Jia X, Durand S, et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 2021, 592: 86–92
 - 37 Yang N, Zeng K, Lu H, et al. MorphGrower: a synchronized layer-by-layer growing approach for plausible neuronal morphology generation. In: *Proceedings of the 41st International Conference on Machine Learning*, 2024
 - 38 Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry. In: *Proceedings of International Conference on Machine Learning*, 2017. 1263–1272
 - 39 Anderson E, Veith G D, Weininger D. SMILES, a line notation and computerized interpreter for chemical structures. US Environmental Protection Agency, Environmental Research Laboratory, 1987. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NHEERL&dirEntryId=33186
 - 40 Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. 2018. ArXiv:1810.04805
 - 41 Morgan H L. The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. *J Chem Doc*, 1965, 5: 107–113
 - 42 Jaeger S, Fulle S, Turk S. Mol2vec: unsupervised machine learning approach with chemical intuition. *J Chem Inf Model*, 2018, 58: 27–35
 - 43 Jin W, Coley C, Barzilay R, et al. Predicting organic reaction outcomes with Weisfeiler-Lehman network. In: *Proceedings of Advances in Neural Information Processing Systems*, 2017. 30
 - 44 Ishida S, Miyazaki T, Sugaya Y, et al. Graph neural networks with multiple feature extraction paths for chemical property estimation. *Molecules*, 2021, 26: 3125
 - 45 Yang N, Zeng K, Wu Q, et al. MoleRec: combinatorial drug recommendation with substructure-aware molecular representation learning. In: *Proceedings of the ACM Web Conference*, 2023. 4075–4085
 - 46 Yan G, Wu H, Yan J. Quantum 3D graph learning with applications to molecule embedding. In: *Proceedings of International Conference on Machine Learning*, 2023. 39126–39137
 - 47 Ma T, Xiao C, Zhou J, et al. Drug similarity integration through attentive multi-view graph auto-encoders. In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018. 3477–3483
 - 48 Wang Y, Min Y, Chen X, et al. Multi-view graph contrastive representation learning for drug-drug interaction prediction. In: *Proceedings of the Web Conference*, 2021. 2921–2933
 - 49 Celebi R, Uyar H, Yasar E, et al. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. *BMC Bioinf*, 2019, 20: 1–4
 - 50 Zhou Y, Qiao Z, Du Y, et al. A survey on the construction methods and applications of sci-tech big data knowledge graph. *Sci Sin Inform*, 2020, 50: 957–987
 - 51 Deac A, Huang Y H, Veličković P, et al. Drug-drug adverse effect prediction with graph co-attention. 2019. ArXiv:1905.00534
 - 52 Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data. In: *Proceedings of Advances in Neural Information Processing Systems*, 2013. 26
 - 53 Landrum G. RDKit: a software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*, 2013, 8: 31
 - 54 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Proceedings of Advances in Neural Information Processing Systems*, 2017. 30
 - 55 Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. In: *Proceedings of International Conference on Learning Representations*, 2018
 - 56 Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. In: *Proceedings of*

- International Conference on Learning Representations, 2016
- 57 Xu K, Hu W, Leskovec J, et al. How powerful are graph neural networks? In: Proceedings of International Conference on Learning Representations, 2018
- 58 Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2014
- 59 Wishart D S, Feunang Y D, Guo A C, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*, 2018, 46: 1074–1082
- 60 Tatonetti N P, Ye P P, Daneshjou R, et al. Data-driven prediction of drug effects and interactions. *Sci Transl Med*, 2012, 4: 125ra31
- 61 Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model*, 2010, 50: 742–754
- 62 Mallick K, Bandyopadhyay S, Chakraborty S, et al. Topo2Vec: a novel node embedding generation based on network topology for link prediction. *IEEE Trans Comput Soc Syst*, 2019, 6: 1306–1317
- 63 Daud N N, Hamid S H A, Saadoon M, et al. Applications of link prediction in social networks: a review. *J Netw Comput Appl*, 2020, 166: 102716
- 64 Gao Z, Jiang C, Zhang J, et al. Hierarchical graph learning for protein-protein interaction. *Nat Commun*, 2023, 14: 1093
- 65 de Meo P, Ferrara E, Fiumara G, et al. Generalized Louvain method for community detection in large networks. In: Proceedings of the 11th International Conference on Intelligent Systems Design and Applications, 2011. 88–93

Drug-drug interaction prediction via hierarchical structure modeling

Huaijin WU¹, Jiazi BU¹, Nianzu YANG¹, Yao SUN², Haitao SONG^{1,3}, Ning LIU¹ & Junchi YAN^{1,4*}

1. *MoE Key Lab of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai 200240, China*

2. *Zhejiang Lab, Hangzhou 311121, China*

3. *Shanghai Artificial Intelligence Research Institute, Shanghai 200240, China*

4. *School of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai 200240, China*

* Corresponding author. E-mail: yanjunchi@sjtu.edu.cn

Abstract Drug-drug interactions (DDIs) often occur when a drug is co-administered with other drugs. Some interactions between drugs can result in positive therapeutic effects, while reactions caused by others can have serious consequences, even life-threatening, making it crucial to predict whether the interaction or what type exists between two drugs. The escalating demand and cost of experimental DDI studies underscore the requirement for automated computational tools, especially machine learning methods, to predict DDIs. However, previous machine learning methods have failed to comprehensively model multiple levels of information, including the substructure level, molecule level, and DDI network level. Some methods also rely on large-scale biomedical knowledge graphs, increasing the inference cost. In this paper, we propose HierDDI, which constructs a hierarchical graph structure solely from known DDIs, without relying on external knowledge graphs. Specifically, HierDDI predicts DDIs through modeling the aforementioned three levels, capturing both inter-drug and intra-drug interactions to establish a comprehensive machine understanding of DDIs. While all modules were designed with simplicity, extensive experiments showcase that our method outperforms baselines by a notable margin. Particularly, on DrugBank and Twosides public benchmarks, our method achieves 0.58% and 2.02% improvement in accuracy over the strongest baselines in the transductive setting, while these two values reach up to 4.20% and 3.52% in the inductive setting. The source code is publicly available at <https://github.com/wuhuaijin/HierDDI>.

Keywords machine learning, drug-drug interaction, hierarchical, graph neural network, end-to-end